ANL-6730

ANL-6730

Argonne National Laboratory

THE MORSE INDEX THEOREM AND GEOMETRICAL OPTICS

by

Robert Hermann

LEGAL NOTICE

ANL-6730
Physics
(TID-4500, 21st Ed.)
AEC Research and
Development Report

ARGONNE NATIONAL LABORATORY
9700 South Cass Avenue
Argonne, Illinois

THE MORSE INDEX THEOREM AND GEOMETRICAL OPTICS

by

Robert Hermann
Northwestern University

Applied Mathematics Division

May 1963

Operated by The University of Chicago
under
Contract W-31-109-eng-38
with the
U. S. Atomic Energy Commission

THE MORSE INDEX THEOREM AND GEOMETRICAL OPTICS

by

Robert Hermann

1. INTRODUCTION

Consider a differential equation of the form

$$v''(t) + r(t)v(t) = 0 \quad . \tag{1.1}$$

Classical Sturm-Liouville theory deals with such equations in which v(t) and r(t) are scalar-valued functions of t[6]. The theory of these equations is well-known, and, of course, they appear in many contexts in both engineering and physics. However, the theory of systems of type (1.1) in which v(t) is a vector-valued function of t is considerably less developed and well-known among workers in applied mathematics, despite the fact that many physical problems lead to such equations in a very natural way, particularly in stability problems. M. Morse has developed the foundations for a successful generalization of the classical Sturm-Liouville theory to such systems [7]. This work has in recent years been extended and applied to various problems in global Riemannian geometry, and has led to a revitalization of this subject. Since this work does not seem to be well-known among its potential consumers in the applied fields, we aim to give here an exposition of some parts. We will assume that the reader is conversant with the basic existence and uniqueness theorems for ordinary differential equations and with the treatment of linear algebra via the theory of vector spaces.

We will now present enough notations and definitions to be able to state the main result: the Morse index theorem. The proof will be given in the second section. Since it may be difficult for the reader to see the forest for the trees while reading the proof, we may point out here that the proof basically consists in putting together certain well-known analytical techniques concerning systems of second-order linear differential equations with the basic ideas of the calculus of variations. The main difference in our proof from Morse's is that we try to work directly with the infinite-dimensional linear spaces that occur, whereas Morse, by a variety of ingenious analytical and geometric tricks, tries to reduce the infinite-dimensional situation to a finite one.

Let V be a vector space of finite dimension[1] over the real numbers. Elements of V will be denoted by such letters as u, v, w, ... It will be assumed that V has a given fixed, positive-definite, symmetric bilinear form $(u,v) \to \langle u,v\rangle$. Thus:

$$\langle au + bv, a_1u_1 + b_1v_1\rangle = aa_1 \langle u,u_1\rangle + ab_1 \langle u,v_1\rangle + a_1b \langle v,u_1\rangle + bb_1 \langle v,v_1\rangle \tag{1.1a}$$

for $a,a_1,b,b_1 \in R$ (= real numbers), $u,v,u_1,v_1 \in V$.

$$\langle u,v\rangle = \langle v,u\rangle \text{ for } u,v \in V \quad . \tag{1.1b}$$

$$\langle u,u\rangle \geq 0 \text{ for } u \in V \quad . \quad \langle u,u\rangle = 0 \text{ if and only if } u = 0 \quad . \tag{1.1c}$$

For $v \in V$, put $\|v\| = \langle v,v\rangle^{1/2}$. Recall that the following inequalities follow from the positive-definite condition:

$\|u + v\| \leq \|u\| + \|v\|$ (triangle inequality)

$|\langle u,v\rangle| \leq \|u\| \; \|v\|$. Equality holds if and only if $au + bv = 0$ for some $a,b \in R$. (Schwarz inequality).

We must also consider linear transformations of V onto itself, usually denoted by R,S,T,..., and bilinear symmetric forms other than $\langle\ ,\ \rangle$ that will not necessarily be positive-definite [i.e., satisfy (1.1a), b), but not c)], and that will be denoted by $Q(\ ,\ )$. A linear transformation $R\colon V \to V$ is said to be symmetric if

$$\langle R(u),v\rangle = \langle u,R(v)\rangle \quad \text{for } u, v \in V \quad .$$

t will be a real parameter extending over the interval $[0, \infty)$ or a subinterval. We will consider vector-valued functions of t, denoted usually by u(t), v(t), etc., defined over an interval and usually continuous, piecewise C^2, and taking values in V. $u'(t)$, $u''(t)$, dv/dt, etc., denote the derivative with respect to t.

We will be considering differential operators of the form

$$v \to v''(t) + R_t(v(t)), \text{ also denoted by:} \tag{1.2}$$

$$J = \frac{d^2}{dt^2} + R_t \quad ,$$

[1] It seems to be an open problem to extend the theory to infinite-dimensional space.

where $t \to R_t$ is a one-parameter family of symmetric linear transformations of V. (It is possible to generalize the theory by including some kinds of terms in v' on the right-hand side of (1.2), but we prefer to treat this simpler case, referring to Morse [7] for a complete treatment.

We must also consider boundary conditions: Algebraically, a boundary condition is an ordered pair (W,Q) consisting of a subspace $W \subset V$ and a bilinear, symmetric form $(u,v) \to Q(u,v)$ defined on W alone.

One fundamental problem may be described as follows: Find a solution of

$$v''(t) + R_t(v(t)) = 0, \quad 0 \le t \le \infty, \quad \text{subject to the following} \tag{1.3}$$

boundary conditions;

$$v(0) \in W, \quad \langle v'(0), w\rangle = -Q(v(0),w) \quad \text{for all } w \in W; \tag{1.4}$$

$$v(a) \in W^a, \quad \langle v'(0), w\rangle = -Q^a(v(a),w) \quad \text{for all } w \in W^a, \tag{1.5}$$

for a given number $a > 0$, and two sets (W,Q) and (W^a,Q^a) of boundary conditions. We refer to (1.4) and (1.5) as, respectively, left- and right-hand boundary conditions.

There is a problem in the calculus of variations associated with (1.3)-(1.5) that is the foundation for the Morse treatment. Proceed as follows to find it:

Suppose $v(t)$, $0 \le t \le a$, satisfies (1.3)-(1.5). Then:

$$-\int_0^a \langle v''(t) + R_t(v(t)), v(t)\rangle$$

$$= -\langle v'(t),v(t)\rangle\Big|_0^a + \int_0^a [\langle v'(t),v'(t)\rangle - \langle R_t(v(t),v(t)\rangle]\,dt$$

$$= Q^a(v(a),v(a)) - Q(v(0),v(0) + \int_0^a [\,\|v'\|^2 - \langle R_t v,v\rangle]\,dt \quad .^1$$

This suggests the following definition: Suppose $v(t)$, $0 \le t \le a$, is a curve in V. Define:

$$I(v) = Q^2(v(a),v(a)) - Q(v(0),v(0)) + \int_0^a \|v'\|^2 - \langle Rv,v\rangle\,dt \quad , \tag{1.6}$$

[1] Where it is felt that it will lead to no confusion, we will compress the notation by omitting t.

and call it the index of the curve v. If our boundary-value problem (1.3-5) admits a solution, there is a curve v with $I(v) = 0$; hence, it is suggested that we turn this remark around and try to minimize $I(v)$ by a curve $v(t)$ satisfying (1.3-5). This is an ordinary variational problem. It is readily verified that its Euler equations are (1.3), but this fact will remain in the background.

In this paper we will restrict ourselves to the case in which the right-hand boundary conditions (W^a, Q^a) are identically zero.

Definition. A point $a \in (0, \infty)$ is said to be a focal point for the operator and boundary condition (W,Q) if there is a non-trivial, C^2 curve $v(t)$ in V, $0 \le t \le a$, satisying:

$$J(v) = 0 \quad ;$$

$$v(0) \in W \quad ;$$

$$\langle v'(0), w\rangle = -Q(v(0),w) \text{ for all } w \in W; \text{ and}$$

$$v(a) = 0 \quad .$$

The index of such a focal point is equal to the dimension of the linear space of all curves satisfying these conditions (hence, is always infinite and no greater than the dimension of V).

Definition. Let $[0,a]$ be an interval of real numbers. Let $\Omega(0,a)$ be the space of continuous, piecewise C^2 curves $t \to v(t)$, $0 \le t \le a$, in V satisfying the following conditions:

$$v(0) \in W \quad ;$$

$$\langle v'(0), w\rangle = -Q(v(0),w) \text{ for all } w \in W \quad ;$$

$$v(a) = 0 \quad .$$

Since two such curves can be added pointwise and multiplied by real constants, $\Omega(0,a)$ is a vector space over the real numbers. For $v \in \Omega(0,a)$, let:

$$I(v) = -Q(v(0),v(0)) + \int_0^a \|v'(t)\|^2 - \langle R_t(v), v\rangle \, dt \quad .$$

Define the index of the interval $[0,a]$ as the maximum number of linearly independent elements of $\Omega(0,a)$ on which the function I is negative.

Thus, there are the two distinct ideas of index of a focal point and index of a closed interval $[0,a]$. They are related via the following main theorem:

Morse Index Theorem. The index of an interval $[0,a]$ is finite and equal to the sum of indices of the focal points contained in the open interval $(0,a)$. It is also equal to the maximal number of linearly independent elements of $\Omega(0,a)$ that are C^2 and are eigenfunctions of the differential operator $J = (d^2/dt^2) + R_t$ for positive eigenvalues.

As a general intuitive remark, notice that the index of an interval is an analytical invariant of the operator J and boundary condition (W,Q), while the sum of the indices of the focal points is more like a topological invariant. Thus, the index of the interval may be expected to vary reasonably smoothly when J,(W,Q), or $[0,a]$ are varied in a reasonably smooth way. As such a variation is performed, it is not expected that each focal point varies smoothly; the remarkable fact contained in the Index Theorem is that the sum of indices of the focal points does vary in a more reasonable way. Another intuitive remark is that the Index Theorem provides the foundation for a perturbation-theory approach to the problem of finding focal points.

2. PROOF OF THE MORSE INDEX THEOREM

Let V, (W,Q), $J = (d^2/dt^2) + R_t$, $\Omega(0,a)$, etc., be as described in the introduction. They will be considered as fixed throughout the discussion.

The proof of the Index Theorem will be broken up into smaller steps.

Lemma 2.1. Given a pair (v_0,u_0) of vectors in V, there is a unique curve in V: $t \to v(t)$, $0 < t < \infty$, satisfying $J(v) = 0$ and: $v(0) = v_0$, $v'(0) = u_0$. In particular, if $v_0 = u_0 = 0$, then $v(t) \equiv 0$. If v_0, u_0, and the coefficients of J depend continuously on additional parameters, so do the resulting solutions, and the dependence is uniformly continuous for t ranging over a bounded closed interval.

This follows from the basic existence theorem for ordinary differential equations [4].

Lemma 2.2. The vector space of solution curves of $J = 0$ that are C^2 and satisfy the (W,Q) boundary condition at $t = 0$ has the same dimension as V.

Proof. For later reference, we will prove a little more and develop additional notations. Suppose $\dim V = n$, $\dim W = m$, $\dim W^\perp = n-m$. ($W^\perp$ denotes the orthogonal complement of W in V with respect to the form $\langle\, ,\rangle$. Explicitly,

$$W^\perp = \{u \in V:\ \langle u,w\rangle = 0 \text{ for all } w \in W\} \quad .)$$

Adopt the following ranges of indices:

$$1 \leq i,j,\ldots \leq n \quad ; \quad 1 \leq a,b,\ldots < m \quad ; \quad m+1 \leq \alpha,\beta,\ldots \leq n \quad . \tag{2.1}$$

Adopt a fixed orthonormal basis (u_i) of V such that (u_a) and (u_α) are, respectively, orthonormal bases of W and $W^\perp$. Then, we can find n-solution curves of $J = 0$, denoted by $v_i(t)$, $0 \leq t < \infty$, $1 \leq i \leq n$, such that:

$$v_a(0) = u_a \quad , \quad 1 \leq a \leq m \quad . \tag{2.2a}$$

$$\left.\begin{array}{l} \langle v_a'(0), w\rangle = -Q(u_a, w) \\ v_a'(0) \in W \end{array}\right\} \text{ for all } w \in W,\ 1 \leq a \leq m \tag{2.2b}$$

$$\left.\begin{array}{ll} v_\alpha(0) = 0 & (2.2c) \\ v_\alpha'(0) = u_\alpha & (2.2d) \end{array}\right\} \text{ for } m+1 \leq \alpha \leq n \quad .$$

(The existence and uniqueness of solution curves satisfying these conditions follow easily from Lemma 2.1.) Note also that these curves satisfy the (W,Q)-boundary condition at $t = 0$.

We show that the curves $v_i(t)$, $1 \leq i \leq n$, are linearly independent. Suppose there is a linear relation of the form:

$$\sum_i^n C_i v_i(t) = 0 \quad .$$

Setting $t = 0$, using (2.2a) and (2.2c), we have:

$$\sum_a C_a u_a = 0 \quad ,$$

implying $C_a = 0$, implying

$$\sum_\alpha C_\alpha v(t) = 0 \quad .$$

Differentiating, setting $t = 0$, and using (2.2d), we have:

$$\sum_\alpha C_\alpha u_\alpha = 0 \quad ,$$

forcing $C_\alpha = 0$, whence linear independence of the $v_i(t)$.

To complete the proof of Lemma 2.2, we show that every solution curve v(t) if J = 0 satisfying the (W,Q)-boundary condition at t = 0 can be written as a sum of the $v_i(t)$ with constant coefficients. Now, we have:

$v(0) \in W$; hence v(0) can be written as:

$$v(0) = \sum_a C_a u_a = \sum_a C_a v_a \; C_a v_a(0)$$

The solution curve $v(t) - \sum_a C_a v_a(t)$ is zero for t = 0; hence its derivative at t = 0 can be written as a sum

$$\sum C_\alpha u_\alpha = \sum_\alpha C_\alpha v'_\alpha(0) \quad .$$

Thus, $v(t) - \sum_i C_i v_i(t)$ is a solution of J = 0, is zero at t = 0, and its first derivative is zero at t = 0; hence is identically zero. Q.E.D.

For future reference, we shall refer to the basis $(v_i(t))$ of solutions of J = 0 and the (W,Q)-boundary condition constructed above as a canonical basis.

Lemma 2.3. *If v(t) and w(t) are two solutions of J = 0 satisfying the (W,Q)-boundary condition at t = 0, then:*

$$\langle v'(t), w(t) \rangle = \langle v(t), w'(t) \rangle \quad \text{for } 0 \le t < \infty \quad . \tag{2.3}$$

Proof. Note the identity:

$$\frac{d}{dt}(\langle v'(t),w(t)\rangle - \langle v(t),w'(t)\rangle) = \langle v''(t),w(t)\rangle + \langle v'(t),w'(t)\rangle - \langle v'(t),w'(t)\rangle - \langle v(t),w''(t)\rangle$$

$$= \langle -R_t(v(t)), w(t) \rangle + \langle v(t), R_t(w(t)) \rangle = 0 \quad ,$$

obtained by use of the symmetry property of R_t. Now,

$$\langle v'(0),w(0)\rangle - \langle v(0),w'(0)\rangle = -Q(v(0),w(0)) + Q(v(0),w(0)) = 0 \quad .$$

Lemma 2.4. *If ϵ is sufficiently small, there are no focal points on the interval $[0,\epsilon]$.*

<u>Proof</u>. Let $(v_i(t))$, $1 < i \leq n$, be a canonical basis for solutions of $J = 0$ satisfying the (W,Q)-boundary condition at $t = 0$.

Define curves $w_i(t)$ as follows:

$$w_a(t) = v_a(t) \text{ for } 1 \leq a \leq m \quad ;$$

$$w_\alpha(t) = \frac{v_\alpha(t)}{t} \text{ for } m+1 \leq \alpha \leq n \quad .$$

By (2.2), w_α is continuous at $t = 0$ and equals there the $v_\alpha'(0) = u_\alpha$. Then the vectors $(w_i(t))$ are linearly independent for $t = 0$, hence by continuity also for t sufficiently small, say for $0 \leq t \leq \epsilon$. Then $[0,\epsilon]$ can contain no focal points. For, suppose otherwise, i.e., $v(t)$ is a solution of $J = 0$ satisfying the (W,Q)-boundary condition at $t = 0$ and vanishing at, say, $t = \epsilon$. By Lemma 2.2, $v(t)$ can be written as $\sum_i C_i v_i(t)$, for constants C_i. Hence, also,

$$0 = v(\epsilon) = \sum_a C_a w_a(\epsilon) + \sum_a C_\alpha w_\alpha(\epsilon)\cdot\epsilon \quad .$$

Thus, $C_a = 0 = C_\alpha$, hence $v(t) \equiv 0$, a contradiction.

<u>Lemma 2.5</u>. <u>Suppose $t_0 \in (0,\infty)$ is a focal point. Then, for ϵ sufficiently small, $[t_0 - \epsilon, t_0 + \epsilon]$ contains no other focal point. We then conclude, using also Lemma 2.4, that each bounded interval contains only a finite number of focal points.</u>

<u>Proof.</u> Suppose $v_i(t)$, $1 \leq i \leq n$, is any basis of solutions of $J = 0$ satisfying the (W,Q)-boundary condition at $t = 0$, such that:

$$v_i(t_0) = 0 \text{ for } 1 \leq i \leq p \quad ,$$

but $v_i(t_0)$ are linearly independent for $p+1 \leq i \leq n$ (p is then the index of the focal point). By formula (2.3),

$$\langle v_i'(t_0), v_j(t_0)\rangle = 0 \text{ for } 1 \leq i \leq p, \text{ and } p+1 \leq j \leq n \quad .$$

The $v_i'(t_0)$ must be linearly independent for $1 \leq i \leq p$ [otherwise, the $v_1(t),\ldots, v_p(t)$ could not be linearly independent]; hence, $v_i'(t_0),\ldots,v_p'(t_0),v_{p+1}(t_0),\ldots, v_n(t_0)$ must form a basis for V. Now,

$$\lim_{t\to t_0} \frac{v_i(t)}{t - t_0} = v_i'(t_0) \text{ for } 1 \leq i \leq p \quad ;$$

hence, if ϵ is sufficiently small, the vectors

$$\frac{v_i(t_0+\epsilon)}{\epsilon} \quad , \quad v_j(t_0+\epsilon), \quad \text{for } 1 \le i \le p \quad , \quad p+1 \le j \le n \quad ,$$

form a basis for V. The proof that there are no focal points on $[t_0-\epsilon,\ t_0+\epsilon]$ is now similar to that in Lemma 2.4.

We need more notation. If $v(t)$ is a curve in $\Omega(0,a)$ and $W(t)$, $0 \le t \le a$, is any continuous piecewise C^1 curve in V satisfying the (W,Q)-boundary condition at $t = 0$, put:

$$I(v,w) = -Q(v(0),w(0)) + \int_0^a \langle v'(t),w'(t)\rangle - \langle R_t(v(t)),w(t)\rangle \, dt \ . \tag{2.4}$$

Let $\Omega_J(0,a)$ be the subset of curves $v(t)$ in $\Omega(0,a)$ defined by taking all linear combinations with constant coefficients of curves of the following type: For each $t_0 \in (0,a]$ that is a focal point, consider a C^2 curve $v(t)$ in $[0,t_0]$ that satisfies $J = 0$ and the (W,Q)-boundary condition at $t = 0$, and that vanishes at $t = t_0$. Extend this curve over $[0,a]$ by defining $v(t) = 0$ for $t_0 \le t \le a$. Graphically,

Then,

The dimension of $\Omega_J(0,a)$ as a real vector space is equal to the sum of indices of the focal points on the interval $(0,a]$. (2.5)

Lemma 2.6. Let $(v_i(t))$, $1 \le i \le n$, $0 \le t \le a$ be any basis for the vector space of curves in V that are C^2 and satisfy $J = 0$. Suppose that $v(t)$, $0 \le t \le a$, is a differentiable curve in V such that:

$I(v,w) = 0$ for all curves $w(t)$, $0 \le t \le a$, that lie in $\Omega_J(0,1)$. Then, $v(t)$ admits a representation as:

$$v(t) = \sum_{i=1}^{n} a_i(t) v_i(t) \quad \text{for } 0 \le t \le a \quad , \tag{2.6}$$

where the coefficients $a_i(t)$ are continuous, piecewise C^2 functions for $0 \le t \le a$.

Proof. Obviously, $v(t)$ admits such a representation (2.6) valid except possibly for the values of t that are focal points. We must show the functions $a_i(t)$ obtained in this way have a limit as t approaches a focal point. Suppose, then, that $t_0 \in (0,a]$ is such a focal point. We may suppose the basis $(v_i(t))$ is chosen so that

$$v_i(t_0) = 0 \quad \text{for } 1 \leq i \leq p \quad ;$$

$$(v_i(t_0)) \text{ are linearly independent for } p+1 \leq i \leq n \quad .$$

By Lemma 2.3, $\langle v_i'(t_0), v_j(t_0)\rangle = 0$ for $1 \leq i \leq p$, $p+1 \leq j \leq n$. As before, this implies that $v_1'(t_0),\ldots,v_p'(t_0)$, $v_{p+1}(t_0),\ldots,v_n(t_0)$ forms a basis for V. Then, for t close to t_0,

$$v(t) = \sum_{i=1}^{p} a_i(t)(t-t_0)\frac{v_i(t)}{t-t_0} + \sum_{i=p+1}^{n} a_i(t)v_i(t) \quad .$$

For $1 \leq i \leq p$,

$$\lim_{t\to t_0} \frac{v_i(t)}{t-t_0} = v_i'(t_0) \quad ;$$

hence for t sufficiently close to t_0,

$$\frac{v_1(t)}{t-t_0},\ldots,\frac{v_p(t)}{t-t_0}, v_{p+1}(t),\ldots,v_n(t)$$

forms a basis of V and depends in a C^1 way on t. Since $v(t)$ is continuous, the functions $a_i(t)(t-t_0)$ for $1 \leq i \leq p$ and $a_i(t)$ for $p+1 \leq i \leq n$ are continuous at t_0.

These remarks are valid for any v that is merely continuous. Now we want to take into account the fact that $I(v,w) = 0$ for all $w \in \Omega_J(0,1)$. For $1 \leq j \leq p$, let w_i be the elements of $\Omega_J(0,1)$ defined as follows:

$$\begin{aligned} w_i(t) &= v_i(t) \quad \text{for } 0 \leq t \leq t_0 \\ w_i(t) &= 0 \quad \text{for } t_0 \leq t \leq a \end{aligned} \qquad (2.7)$$

$$. \; 0 = I(v,w_i) = -Q(v(0),w_i(0)) + \int_0^a \langle v'(t),w_i'(t)\rangle - \langle R_t(v(t)),w_i(t)\rangle \, dt$$

equals, using (2.7),

$$-Q(v(0),v_i(0)) + \int_0^{t_0} \langle v'(t),v_j'(t)\rangle - \langle R_t(v(t)),v_j(t)\rangle \, dt$$

equals, after integrating by parts and taking into account the fact that $J(v_i) = 0$ and that v and v_i satisfy the (W,Q)-boundary condition at $t = 0$,

$$\langle v(t_0),\ v_j'(t_0)\rangle \quad .$$

Thus, $v(t_0)$ must be a linear combination of $v_{p+1}(t_0),\ldots,v_n(t_0)$. We conclude that

$$\lim_{t\to t_0} a_i(t)(t-t_0) = 0 \qquad \text{for } 1 \le i \le p \quad .$$

Now, since $v(t)$ is differentiable, the functions $a_i(t)(t-t_0)$ for $1 \le i \le p$ are differentiable at $t = t_0$. We conclude (using the definition of derivative) that $\lim_{t\to t_0} a_i(t)$ exists and equals

$$\frac{d}{dt}\left(a_i(t)(t-t_0)\right)\Big|_{t=t_0} \quad . \qquad \underline{\text{Q.E.D.}}$$

<u>Lemma 2.7.</u> <u>Let $v_i(t)$, $1 \le i \le n$, $0 \le t \le a$, be a basis of curves in V that are C^2 and satisfy $J = 0$ and the (W,Q)-boundary condition at $t = 0$. Suppose $u(t)$ and $v(t)$, $0 \le t \le a$, are two curves in V admitting representations of the following type:</u>

$$u(t) = \sum_{i=1}^{n} f_i(t)v_i(t) \qquad \text{for } 0 \le t \le a$$

$$v(t) = \sum_{i=1}^{n} f_i(a)v_i(t) \qquad \text{for } 0 \le t \le a \quad .$$

<u>Suppose in addition that the functions $f_i(t)$ are continuous and piecewise C^1 for $0 < t \le a$, and that $u(t)$ satisfies the (W,Q)-boundary condition at $t = 0$. Then</u>

$$I(u) \ge I(v) \quad .$$

<u>Equality holds only if $u = v$.</u>

<u>Proof.</u> For $\epsilon > 0$, let

$$I_\epsilon(u) = -Q(u(0),u(0)) + \int_\epsilon^a \|u'(t)\|^2 - \langle R_t(u(t)),u(t)\rangle\, dt \quad .$$

$$u'(t) = \sum_{i=1}^{n} (f_i'(t)v_i(t) + f_i(t)v_i'(t))$$

$$\|u'(t)\|^2 = \sum_{i,j=1}^{n} \Big(f_i'(t)f_j'(t) <v_i(t),v_j(t)> + f_i(t)f_j(t) <v_i'(t),v_j'(t)>$$

$$+ f_i'(t)f_j(t) <v_i(t),v_j'(t)> + f_i(t)f_j'(t) <v_i'(t),v_j(t)>\Big) \quad .$$

Now,

$$\int_\epsilon^a f_i(t)f_j(t) <v_i'(t),v_j'(t)> dt = f_i(t)f_j(t) <v_i'(t),v_j(t)> \Big|_{t=\epsilon}^{a}$$

$$- \int_\epsilon^a f_i'(t)f_j(t) <v_i'(t),v_j(t)> + f_i(t)f_j'(t) <v_i'(t),v_j(t)>$$

$$+ f_i(t)f_j(t) <v_i''(t),v_j(t)> dt \quad .$$

Using the last two identities and Lemma 2.3, we have

$$\int_\epsilon^a \|u'(t)\|^2 dt = \int_\epsilon^a \Big[\|\sum_{i=1}^{n} f_i'(t)v_i(t)\|^2 + <R_t(u(t)), u(t)> \Big] dt$$

$$- Q(u(a),u(a)) + Q(u(\epsilon),u(\epsilon)) \quad .$$

Hence:

$$I_\epsilon(u) = \int_\epsilon^a \|\sum_{i=1}^{n} f_i'(t)v_i(t)\|^2 dt$$

$$- Q(u(a),u(a)) + \Big(Q(u(\epsilon),u(\epsilon)) - Q(u(0),u(0))\Big) \quad .$$

Similarly,

$$I_\epsilon(v) = -Q(v(a),v(a)) + (Q(v(\epsilon),v(\epsilon)) - Q(u(0),u(0)) \quad ;$$

hence

$$I_\epsilon(u)-I_\epsilon(v) = \int_\epsilon^a \|\sum_{i=1}^{n} f_i'(t)v_i(t)\|^2 dt$$

$$+ (Q(u(\epsilon),u(\epsilon)) - Q(u(0),u(0)) - \Big(Q(v(\epsilon),v(\epsilon)) - Q(v(0),v(0))\Big) \quad .$$

Since all the other terms in this identity approach a limit, we have:

$$\lim_{\epsilon \to 0} \int_{\epsilon}^{a} \left\| \sum_{i=1}^{n} f_i'(t) v_i(t) \right\|^2 dt \text{ exists.}$$

Since it must clearly be >0 unless $f_i'(t) = 0$, i.e., unless $u(t) = v(t)$ for $0 \le t \le a$, we have

$$I(u) > I(v), \quad \text{except if } u = v. \qquad \underline{\text{Q.E.D.}}$$

This lemma is due to W. Ambrose [2,3] and serves as a replacement for the arguments from the general calculus of variations that were used by Morse.

Corollary to Lemma 2.7. The interval $[0,a]$ contains no focal points if and only if $I(u) > 0$ for all curves $u \in \Omega(0,a)$. (In other words, the Morse index theorem holds if $[0,a]$ contains no focal point.)

We must now apply Lemma 2.7 in the special case in which W and Q are both zero; focal points are, in this case, called conjugate points. For the reader's convenience, we restate the definition in slightly different form.

Definition. Let a and b be positive real numbers; a and b are said to be mutually conjugate if there is a C^2 curve $v(t)$, not identically zero, satisfying:

$$v'' + R_t(v(t)) = 0 \quad ,$$

and

$$v(a) = v(b) = 0 \quad .$$

Lemma 2.8. Suppose that a and b are real numbers, $0 \le a < b$, such that the real number interval between them contains no pair of mutually conjugate points. Suppose that $u(t)$ and $v(t)$ $0 \le t < \infty$, are continuous curves such that:

u is piecewise C^2 and v is C^2.

v satisfies: $\left(\frac{d^2}{dt^2} + R_t\right)(v) = 0.$

$u(b) = v(b)$

$u(a) = v(a).$

Then,

$$\int_a^b \|v'(t)\|^2 - \langle R_t(v(t)),v(t)\rangle\, dt \le \int_a^b \|u'(t)\|^2 - \langle R_t(u(t)),u(t)\rangle\, dt \quad .$$

Equality holds only if $u(t) = v(t)$ for $a \le t \le b$.

Proof. First we deal with the case $u(a) = v(a) = 0$. By a translation of the origin of the t-axis, we can also suppose that $a = 0$. The result then follows from Lemma 2.7, since our hypotheses imply that there are no focal points in the interval $(0,b]$, with respect to the boundary conditions $W = 0$, $Q = 0$, at $t = 0$.

Now we reduce the general case $u(a) = v(a)$ to the case just considered. By Lemma 2.2 (since a and b are not mutually conjugate), there is a C^2 curve $w(t)$ satisfying $w(b) = 0$, $w(a) = u(a) = v(a)$,

$$\left(\frac{d^2}{dt^2} - R_t\right) w = 0 \quad .$$

Let $u^*(t) = u(t) - w(t)$, $v^*(t) = v(t) - w(t)$. Since $u^*(a) = 0 = v^*(a)$, case 1 applies, to give

$$\int_a^b \|u^{*\prime}(t)\|^2 - \langle R_t(u^*(t)),u^*(t)\rangle\, dt \ge \int_a^b \|v^{*\prime}(t)\|^2$$

$$- \langle R_t(v^{*\prime}(t)),(v^*(t))\rangle\, dt \quad .$$

But, the left-hand side of this inequality is:

$$\int_a^b \Big[\|u'(t)\|^2 + \|w'(t)\|^2 - 2\,\langle u'(t),w'(t)\rangle - \langle R_t(u(t)),u(t)\rangle$$

$$- \langle R_t(w(t)),w(t)\rangle + 2\,\langle R_t(u(t)),w(t)\rangle\Big]\, dt$$

equals, after integrating by parts and taking into account the relations satisfied by w,

$$\int_a^b \Big[\|u'(t)\|^2 - \langle R_t(u(t)),u(t)\rangle\Big]\, dt - \langle w'(a),w(a)\rangle$$

$$+ 2\,\langle u(a),w'(a)\rangle - 2\,\langle u(b),w'(b)\rangle = \int_a^b \Big[\|u'(t)\|^2$$

$$- \langle R_t(u(t)),u(t)\rangle\Big]\, dt + \langle v(a),w'(a)\rangle - 2\,\langle v(b),w'(b)\rangle$$

[since $v(a) = u(a) = w(a)$] .

Now,

$$\left(\frac{d^2}{dt^2} - R_t\right) v^* = 0 \quad ;$$

hence, the right-hand side of the above inequality is, after an integration by parts,

$$<v^{*\prime}(b), v^*(b)> - <v^*(a), v^{*\prime}(0)> = <v'(b) - w'(b), v(b)> \quad ,$$

since $v^*(a) = 0$, and $w(b) = 0$. Similarly,

$$\int_a^b \|v'(t)\|^2 - <R_t(v(t)), v(t)> \, dt = <v'(b), v(b)> - <v'(a), v(a)> \quad .$$

Thus,

$$\int_a^b \|u'(t)\|^2 - <R_t(u(t)), u(t)> \, dt \geq 2 <v(b), w'(b)> - <v(a), w'(a)>$$

$$+ <v'(b), v(b)> - <w'(b), v(b)> = <v(b), w'(b)> - <v(a), w'(a)>$$

$$+ \int_a^b \|v'(t)\|^2 - <R_t(v(t)), v(t)> \, dt + <v'(a), v(a)> \quad .$$

Now,

$$\frac{d}{dt}(<v(t), w'(t)> - <v'(t), w(t)>)$$

$$= <v'(t), w'(t)> + <v(t), w''(t)> - <v''(t), w(t)> - <v'(t), w'(t)>$$

$$= <v(t), R_t(w'(t))> - <R_t(v(t)), w(t)> = 0 \quad .$$

Hence,

$$<v(b), w'(b)> = <v'(b), w(b)> + <v'(a), w(a)> - <v(a), w'(a)>$$

equals, after using the relations $w(b) = 0$ and $w(a) = v(a)$,

$$<v'(a), v(a)> - <v(a), w'(a)> \quad .$$

Putting the last relation together with the last inequality proves Lemma 2.8.

Lemmas 2.7 and 2.8 may be regarded as particular analytical tools needed to prove the Index Theorem. They really contain fundamental facts from the calculus of variations in a disguised form.

Now we proceed to another analytical tool, proving, again in slightly disguised form, a "compactness"principle for solutions of the type of differential equations we have been considering.

Theorem 2.9. Suppose that $J^k : (d^2/dt^2) - R_t^k$ is a sequence of differential operators of the type we have been considering, $k = 1,2,\ldots$, that (W^k, Q^k) is a sequence of boundary conditions, and that a_k is a sequence of real numbers.

Suppose that:

$$\lim_{k\to\infty} R_t^k = R_t \quad ;$$

$$\lim_{k\to\infty} W^k = W \quad ; \qquad \lim_{k\to\infty} Q^k = Q \quad ;$$

$$\lim_{k\to\infty} a_k = a > 0 \quad .$$

Suppose further that $v_k(t)$, $0 \leqslant t \leqslant a_k$, is a sequence of C^2 curves in V satisfying:

the (W^k, Q^k)-boundary condition at $t = 0$

$$v_k(a_k) = 0$$

$$J^k(v_k) = 0$$

$$\int_0^{a_k} \|v_k(t)\| \, dt \leq 1 \quad .$$

Then, at least one subsequence of the v_k converges, along with its first two derivatives, uniformly to a C^2 curve $v(t)$, $0 \leq t \leq a$, that is a solution of $J(v) = 0$, the (W,Q)-boundary condition, and $v(a) = 0$.

Proof.

$$\text{Let } u_k(t) = \frac{v_k(t)}{\|v_k(0)\| + \|v_k'(0)\|} \quad .$$

Since $\|u_k(0)\|$ and $\|u_k'(0)\|$ are ≤ 1, we can suppose after at most taking subsequences that $\lim_{k\to\infty} u_k(0) = u_0$, $\lim_{k\to\infty} u_k'(0) = u_1$. By the existence theorem for ordinary linear differential equations of the type $J^k = 0$, $u_k(t)$ converges, along with its first two derivatives, uniformly to a C^2 nonidentically zero curve $u(t)$, $0 \leq t \leq a$, that is a solution of $J(u) = 0$, the (W,Q)-boundary condition, and $u(a) = 0$. Then, also:

$$\lim_{k\to\infty} \int_0^{a_k} \|u_k(t)\|^2\, dt = \int_0^a \|u(t)\|\, dt \quad .$$

But also,

$$\int_0^{a_k} \|u_k(t)\|\, dt \leq \frac{1}{\|v_k(0)\| + \|v_k'(0)\|} \quad .$$

Hence, $\|v_k(0)\| + \|v_k'(0)\|$ itself is bounded, and we can apply the existence theorem for the systems $J^k = 0$ to infer the existence of a curve $v(t)$ towards which $v_k(t)$ and its first two derivatives converge uniformly. Q.E.D.

We can now proceed to the proof of the Index Theorem. It is most convenient to arrange the proof so that the final result will appear as a statement that different kinds of indices are in reality the same. Hence, we now introduce the different indices, and also the so-called augmented indices.

Definition. Let a be a positive real number.

$I_1(0,a)$ = The sum of indices of the focal points contained in the interval $[0,a)$.

$AI_1(0,a)$ = The sum of indices of the focal points contained in the interval $[0,a]$.

(Thus, AI_1 is the augmented index corresponding to the index I_1.)

$I_2(0,a)$ = The maximal dimension of a linear subspace of $\Omega(0,a)$ on which the form $v \to I(v)$ is negative definite.

$AI_2(0,a)$ = The maximal dimension of a linear subspace if $\Omega(0,a)$ on which the form $v \to I(v)$ is negative semidefinite.

$I_3(0,a)$ = The maximal number of linearly independent, C^2 eigenfunctions of the differential operator $(d^2/dt^2) + R_t$ corresponding to positive eigenvalues that also satisfy the boundary conditions implied by membership in $\Omega(0,a)$.

$AI_3(0,a)$ = The maximal number of linearly independent C^2 eigenfunctions of the differential operator $(d^2/dt^2) + R_t$ corresponding to non-negative eigenvalues that also satisfy the boundary conditions implied by membership in $\Omega(0,a)$.

Note that several facts follow readily:

$$AI_3(0,a) - I_3(0,a) = AI_1(0,a) - I_1(0,a) \tag{2.8}$$

= the index of the focal point w (if a is a focal point)

= 0, if a is not a focal point.

$$I_3(0,a) \leq I_2(0,a) \tag{2.9}$$

$$AI_3(0,a) \leq (AI_2(0,a)) \quad .$$

Proof. Suppose that $v(t)$, $0 \leqslant t \leqslant a$, lies in $\Omega(0,a)$ and satisfies:

$$v''(t) + R_t(v(t)) = \lambda^2 v(t) \quad , \quad \text{with } \lambda \geq 0 \quad .$$

Then,

$$-\lambda^2 \int_0^a \|(v(t))\|^2 \, dt = -\int_0^a \langle v(t,v''(t)) + R_t(v(t))\rangle \, dt$$

equals, after integrating by parts and taking into account the boundary conditions satisfied by v,

$$-Q(v(0),v(0)) + \int_0^a \|v'(t)\|^2 - \langle v(t),R_t(v(t))\rangle \, dt = I(v) \quad .$$

Then, $I(v) \leq 0$.

To prove (2.9), notice now that $v \to I(v)$ restricted to the linear subspace of $\Omega(0,a)$ spanned by the positive eigenfunctions of $(d^2/dt^2) + R_t$ is negative definite. Similarly, $v \to I(v)$ is negative semidefinite on the subspace of $\Omega(0,a)$ spanned by the non-negative eigenfunctions of $\Omega(0,a)$.

Lemma 2.10. $AI_3(0,a)$ is finite.

Proof. Suppose otherwise, i.e., there are an infinite number of C^2 eigenfunctions $v_k(t)$, $k=1,2,\ldots$, $0 \leq t \leq 1$, of $(d^2/dt^2) + R_t$ corresponding to eigenvalues λ_k^2 and satisfying the boundary condition corresponding to lying in $\Omega(0,a)$.

We can suppose without loss of generality that

$$|Q(v_k(0), v_k(0))| + \int_0^a \langle v_k(t), v_k(t)\rangle \, dt = 1 \tag{2.10a}$$

$$\int_0^a \langle v_k(t), v_j(t)\rangle \, dt = 0 \quad \text{if } k \neq j \quad . \tag{2.10b}$$

Then,

$$\lambda_k^2 = \int_0^a \langle v_k'' + R_t(v_k(t)), v_k(t)\rangle \, dt = -Q(v_k(0), v_k(0))$$

$$- \int_0^a \|v_k'(t)\|^2 - \langle R_t(v_k(t)), v_k(t)\rangle \, dt \quad .$$

$$\leq -Q(v_k(0), v_k(0)) + \int_0^a \langle R_t(v_k(t)), v_k(t)\rangle \, dt \quad .$$

Now, there is a real number δ such that:

$$\langle R_t(v), v\rangle \leq \delta \langle v, v\rangle \quad \text{for all } v \in V, \quad 0 \leq t \leq a \quad .$$

Then,

$$\lambda_k^2 \leq |Q(v_k(0), v_k(0))| + \delta \int_0^a \langle v_k(t), v_k(t)\rangle \, dt$$

$$\leq 1 + \delta \quad ,$$

i.e., the sequence (λ_k^2) is bounded. We can then suppose, after possibly taking subsequences, that

$$\lim_{k \to \infty} \lambda_k = \lambda \quad ,$$

and that $v_k(t)$ converges as $k \to \infty$, along with its first two derivatives, uniformly to a curve $v(t)$ that is an eigenfunction for eigenvalue λ (using Theorem 2.9). But, this contradicts (2.10b).

Lemma 2.11.

$$AI_3(0,a) = AI_2(0,a) \quad .$$

$$I_3(0,a) = I_2(0,a) \quad .$$

Proof. We prove the first equality: The second is similar. Let $d = AI_3(0,a)$. Then, there are d linearly independent eigenfunctions $v_1(t), \ldots, v_d(t)$ in $\Omega(0,a)$, with eigenvalues $\lambda_1^2, \ldots, \lambda_d^2$. We can normalize so that:

$$\int_0^a <v_j(t), v_k(t)> dt = \delta_{jk} \quad \text{for } 1 \leq j,k \leq d \quad .$$

What we must show is that, if $v(t) \in \Omega(0,a)$ satisfies:

$$\int_0^a <v_k(t), v(t)> = 0 \quad \text{for } 1 \leq k \leq d \quad , \tag{2.11}$$

then $I(v) > 0$. Suppose otherwise, i.e., such a v exists with $I(v) \leq 0$. Now, minimize $I(v)$ over all $v \in \Omega(0,a)$ satisfying (2.11) and $\int_0^a ||v(t)||^2 \, dt = 1$.

Using the "direct method" of the calculus of variations, [1] this minimum is taken on by a C^2 function $v_0(t)$ in $\Omega(0,a)$ that is also an eigenfunction of $(d^2/dt^2) + R_t$ with eigenvalue λ_0. But, this eigenfunction would then have to satisfy:

$$0 \geq I(v_0) = - \lambda_0 \int_0^a ||v_0(t)||^2 \, dt \quad ,$$

forcing $\lambda_0 \geq 0$, contradicting the definition of d.

Lemma 2.12. If $0 \leq a \leq b$, then

$$I_2(0,a) \leq I_2(0,b)$$

$$AI_2(0,a) \leq AI_2(0,b)$$

$$AI_1(0,a) \leq I_2(0,b) \quad .$$

Proof. Choose ϵ sufficiently small and positive so that $a + \epsilon \leq b$, and so that there are no mutually conjugate points on the interval $[a-\epsilon, a+\epsilon]$. We are now going to define a linear mapping

$$\phi_\epsilon\colon\ \Omega(0,a) \to \Omega(0,b)$$

such that:

$$I(\phi_\epsilon(v)) \le I(v) \quad \text{for each } v \in \Omega(0,a) \tag{2.12}$$

by making use of Lemma 2.8. Explicitly, $\phi_\epsilon(v)(t)$, $0 \le t \le b$, is to be a continuous piecewise C^2 curve in V such that:

$$\phi_\epsilon(v)(t) = v(t) \quad \text{for } 0 \le t \le a-\epsilon \tag{2.13a}$$

$$\phi_\epsilon(v)(t) \text{ is a solution of } \frac{d^2}{dt^2} + R_t = 0 \quad , \tag{2.13b}$$

with:

$$\phi_\epsilon(v)(a-\epsilon) = v(a-\epsilon)$$

$$\phi_\epsilon(v)(a+\epsilon) = 0 \quad .$$

$$\phi_\epsilon(v)(t) = 0 \quad \text{for } a+\epsilon \le t \le b \quad . \tag{2.13c}$$

The reader will readily verify that ϕ_ϵ is a bonafide linear mapping that (using Lemma 2.8) satisfies (2.12). Suppose now that $v_1(t),\dots,v_d(t)$ are linearly independent elements of $\Omega(0,a)$ on which the form I is < 0. By (2.12), $I(\phi_\epsilon(v_k)) < 0$ for $1 \le k \le d$. We must then show that $\phi_\epsilon(v_k)(t)$ are linearly independent if ϵ is sufficiently small. Suppose otherwise, i.e., there is, for each ϵ, a relation of the form: $\sum_{k=1}^{d} a_k(\epsilon)v_k(t) = 0$ valid for $0 \le t \le a-\epsilon$.

We can normalize so that

$$\sum_{k=1}^{d} a_k(\epsilon)^2 = 1 \quad .$$

Then, there would be a sequence of ϵ going to zero such that $a_k(\epsilon) \to a_k$, and a relation of the form:

$$\sum_{k=1}^{d} a_k\, v_k(t) = 0$$

valid for $0 \le t \le 1$, contradicting that the v_k were linearly independent.

This proves the first inequality in Lemma 2.12. The second is similar. The third involves a slight modification of the argument. Let $a_1,\ldots,a_f$ be the focal points in the interval $[0,a]$, arranged so that:

$$0 < a_1 < a_2 < \ldots < a_f \leq a \quad .$$

Let $d_1,\ldots,d_f$ be the indices of each of the focal points. Let $v_1(t),\ldots,v_{d_1}(t)$, $0 \leq t \leq a_1$, be in $\Omega(0,a_1)$, be linearly independent, and C^2, and satisfy $(d^2/dt^2) + R_t = 0$. Then, $\phi_\epsilon(v_1),\ldots, \phi_\epsilon(v_{d_1})$ span a subspace of $\Omega(0,b)$ on which I is < 0. To see this, use the criterion for equality in Lemma 2.8 and the fact that $v_k(t)$ cannot be zero for t sufficiently close to a_1, $1 \leq k \leq d_1$. Similarly, apply this construction to each of the focal points. It is easily seen that the subspaces of $\Omega(0,b)$ obtained in this way are all linearly independent of each other, hence span a subspace of $\Omega(0,b)$ of dimension $d_1+\ldots+d_f = AI_1(0,a)$.

Lemma 2.13. If ϵ is sufficiently small,

$$AI_3(0,a+\epsilon) \leq AI_3(0,a) \quad .$$

Proof. Suppose otherwise. Let ϵ_k, $k=1,2,\ldots$, be a sequence of real numbers, with $\epsilon_k \to 0$ as $k \to \infty$, and with

$$AI_3(0,a+\epsilon_k) \geq AI_3(0,a) + 1 \quad .$$

Let $v_{j,k}(t)$, $0 \leq j \leq AI_3(0,a) + 1$, $0 \leq k < \infty$, $0 \leq t \leq a+\epsilon_k$, be curves in $\Omega(0,a+\epsilon_k)$ that are C^2, that are eigenfunctions of $(d^2/dt^2) + R_t$ with non-negative eigenvalues, and that, for fixed k, are linearly independent. Using Theorem 2.9, we can arrange (by taking subsequences and normalizing) that:

$$\lim_{k\to\infty} v_{j,k}(t) = v_j(t) \quad ,$$

$$\int_0^{a+\epsilon_k} \langle v_{j_1,k}(t), v_{j_2,k}(t)\rangle \, dt = \delta_{j_1 j_2} \quad ,$$

that the first two derivatives converge uniformly for $0 \leqslant t \leqslant a$, and that the corresponding eigenvalues of $(d^2/dt^2) + R_t$ converge. Then,

$$\int_0^{a} \langle v_{j_1}(t), v_{j_2}(t)\rangle \, dt = \delta_{j_1 j_2}, \quad \text{for } 1 \leq j_1, j_2 \leq AI_1(0,a) + 1 \quad .$$

But, the v_j are C^2, belong to $\Omega(0,a)$, and are eigenfunctions of $(d^2/dt^2) + R_t$ with non-negative eigenvalues, and are linearly independent, a contradiction.

Now we can prove the Index Theorem itself. Note that it is equivalent to the statement:

$$I_1(0,a) = I_2(0,a) = I_3(0,a) \quad . \tag{2.14}$$

Note that we have already proved the second of these equalities (Lemma 2.11). The corollary to Lemma 2.7 implies that the first holds if a is sufficiently small. Assuming that (2.14) is true for a, we shall prove it is true for $a+\epsilon$, if ϵ is sufficiently small and positive. Now, $I_1(0,a+\epsilon) = I_1(0,a)$ + (index of the focal point a). By Lemma 2.12,

$$I_3(0,a) = I_2(0,a) \le I_2(0,a+\epsilon) = I_3(0,a+\epsilon)$$

$$\le AI_3(0,a) \quad \text{if } \epsilon \text{ is sufficiently small, by Lemma 2.13,}$$

$$= I_3(0,a) + \text{index of a.}$$

$$I_1(0,a+\epsilon) \le I_2(0,a+\epsilon) = I_3(0,a+\epsilon) \le I_3(0,a) + \text{index of a.}$$

Thus, $I_3(0,a+\epsilon)$ - index of a $= I_1(0,a)$, which proves (2.14) for $a+\epsilon$. Then, the set of all numbers b such that (2.14) is true for all $a \in [0,b]$ is open. To complete the proof, we must show that it is closed. Suppose, then, that a_k is a monotone-increasing sequence,

$$\lim_{k\to\infty} a_k = a \quad ,$$

such that (2.14) is true for each a_k. We must prove it is true for a also. Thus, we have:

$$I_1(0,a_k) = I_2(0,a_k) \text{ for } k=1,2,\ldots \quad .$$

From the definition of I_1, we have:

$$\lim_{k\to\infty} I_1(0,a_k) = I_1(0,a) \quad .$$

By Lemma 2.12, $I_2(0,a_k) \le I_2(0,a)$.

Let us suppose that $\lim_{k\to\infty} I_2(0,a_k) \neq I_2(0,a)$; i.e.,

$I_2(0,a) > I_2(0,a_k)$ for all k.

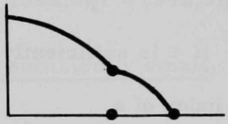
To complete the proof, we will show that:

$$I_3(0,a) \leq I_2(0,a-\epsilon) \quad \text{for } \epsilon \text{ sufficiently small} \quad . \tag{2.15}$$

Suppose, then, that $v_1(t),\dots,v_d(t)$ are linearly independent C^2 eigenfunctions of $(d^2/dt^2) + R_t$ for positive eigenvalues $\lambda_1^2,\dots,\lambda_d^2$ that lie in $\Omega(0,a)$ ($d = I_3(0,a)$). Choose $\epsilon > 0$ sufficiently small, so that there are no mutually conjugate points on the interval $[a,a-2\epsilon]$. Let $v_1^*(t),\dots,v_d^*(t)$, $0 \leq t \leq a-\epsilon$, be continuous piecewise C^2 curves in V defined as follows:

$$v_k^*(t) = v_k(t) \quad \text{for } 0 \leq t \leq a-2\epsilon \quad , \quad 1 \leq k \leq d \quad .$$

$$\frac{d^2}{dt^2} + R_t(v_k^*)(t) = 0 \quad \text{for } a-2\epsilon \leq t \leq a-\epsilon, \quad 1 \leq k \leq d$$



(Uses Lemma 2.8 to construct these curves.) Since $\lim_{\epsilon \to 0} I(v_k^*) = I(v_k) < 0$, for $1 \leq k \leq d$, we see that $I(v_k^*) < 0$ if ϵ is sufficiently small. By an argument similar to that used in proving Lemma 2.12, we see that $v_1^*,\dots,v_k^*$ are linearly independent if ϵ is sufficiently small. This proves (2.14) and finishes the proof of the Index Theorem itself.

3. FOCAL POINTS IN CLASSICAL MECHANICS AND GEOMETRIC OPTICS

We shall take as a starting point the Hamiltonian "model" of classical mechanics [8]. Thus, we are given a space of variables $q_1,\dots,q_n$, describing configuration space, a space of variables $p_1,\dots,p_n$, describing momentum space, and the product space of variables (q_i,p_i), $1 \leq i \leq n$, describing phase space. Where convenient, we will use vector notation: $p = (p_i)$, $q = (q_i)$, etc. An additional variable, labelled as t, is given as the time variable. A real-valued function $H(p,q,t)$ on (phase x time)-space, the Hamiltonian (usually the energy) of the system, describes the development of the system in time via the Hamilton equations:

$$\frac{dq_i}{dt} = \frac{\partial H}{\partial p_i}(q(t),p(t),t) \quad ;$$

$$\frac{dp_i}{dt} = \frac{-\partial H}{\partial q_i}(q(t),p(t),t) \quad . \qquad 1 \leq i \leq n \quad . \tag{3.1}$$

Using a vector notation, we rewrite this as:

$$\frac{dq}{dt} = \frac{\partial H}{\partial p}(q(t),p(t),t) \quad ;$$
$$\frac{dp}{dt} = -\frac{\partial H}{\partial q}(q(t),p(t),t) \quad . \tag{3.1'}$$

Now, suppose that we are given a one-parameter family of solutions of (3.1), depending on, say, a parameter s, $0 \le s \le 1$, reducing, when $s = 0$, to a given solution $(q(t),p(t))$. Thus, analytically, we are given functions $(q(s,t),p(s,t))$ of two variables, with:

$$\frac{\partial q}{\partial t} = \frac{\partial H}{\partial p}(q,p,t) \quad ;$$
$$\frac{\partial p}{\partial t} = -\frac{\partial H}{\partial q}(q,p,t) \quad . \tag{3.2}$$

Suppose that

$$Q(t) = \left.\frac{\partial q}{\partial s}\right|_{s=0} \quad ; \quad P(t) = \left.\frac{\partial p}{\partial s}\right|_{s=0} \quad .$$

Hence, $(q(s,t), p(s,t))$ admits a Taylor expansion of the form:

$$q(s,t) = q(t) + Q(t)s + \ldots \quad ;$$
$$p(s,t) = p(t) + P(t)s + \ldots \quad . \tag{3.3}$$

It is then a reasonable supposition that the behavior of the solution $t \to (q(s,t),p(s,t))$ for nonzero but small s is, compared with $t \to (q(t),p(t))$, regulated by the behavior of the functions $t \to (Q(t),P(t))$. Actually, from the technical point of view things are a good deal more complicated than this, but at any rate in applied problems this is a supposition that can be checked by experiment and that can be used as a basis for design. Now, the functions $(Q(t),P(t))$ themselves are the solution of a system of ordinary differential equations obtained by differentiating (3.2) with respect to s, then setting $s = 0$.

$$\frac{\partial Q}{\partial t} = \frac{\partial^2 H}{\partial p \partial p} P(t) + \frac{\partial^2 H}{\partial q \partial p} Q(t) \quad ;$$
$$\frac{\partial P}{\partial t} = -\frac{\partial^2 H}{\partial p \partial q} P(t) - \frac{\partial^2 H}{\partial q \partial q} Q(t) \quad . \tag{3.4}$$

These are the <u>tensor variational equations</u> of the equations (3.1), based on the given solution $(q(t),p(t))$. The coefficients are functions of t (e.g., $(\partial^2 H/\partial p_i \, \partial p_j)\,(q(t),p(t),t))$ that depend on H and the base solution $(q(t),p(t),t)$.

A curve $(q(t))$ in q-space which is the projection of a curve in phase space of a solution of (3.1) is called an <u>extremal</u>. (In case the Hamiltonian is derived from a problem in the calculus of variations, these are just the curves which satisfy the Euler-Lagrange equations, whence the name.) In physical problems, extremals do not occur individually, but occur imbedded in certain families of extremals, called <u>extremal fields</u>. Each such extremal field is determined by a solution of the <u>Hamilton-Jacobi partial differential equation</u>, which is a real-valued function $S(q,t)$ on q x t-space such that:

$$\frac{\partial S}{\partial t} + H\left(q, \frac{\partial S}{\partial q}, t\right) = 0 \quad . \tag{3.5}$$

In fact, the curves satisfying the following system of ordinary differential equations are extremals, forming an n-parameter family (a "congruence," in classical language) of extremals that will be the extremal field corresponding to $S(q,t)$.

$$\frac{dq}{dt} = \frac{\partial H}{\partial p}\left(q, \frac{\partial S}{\partial q}, t\right) \quad . \tag{3.6}$$

To verify this statement, we will show that the solutions of (3.6) are extremals. Let $p(t)$ be the curve

$$p(t) = \frac{\partial S}{\partial q}\,(q(t),t) \tag{3.7}$$

for a given solution of $q(t)$ of (3.6).

We must show that $(q(t),p(t))$ is a solution of the Hamilton equations (3.1). Now, (3.6) and (3.7) together give one-half of the Hamilton equations. The second half will follow from the fact that $S(q,t)$ is a solution of the Hamilton-Jacobi equation (3.5). In fact:

$$\begin{aligned} \frac{dp}{dt}(t) &= \frac{\partial^2 S}{\partial q\,\partial q}\,(q(t),t)\,\frac{dq}{dt}\,(t) + \frac{\partial^2 S}{\partial q\,\partial t}\,(q(t),t) \\ &= \frac{\partial^2 S}{\partial q\,\partial q}\,(q(t),t)\,\frac{\partial H}{\partial p}\,\left(q(t), \frac{\partial S}{\partial q}(q(t),t\right) + \frac{\partial^2 S}{\partial q\,\partial t}\,(q(t),t) \quad . \end{aligned}$$

From (3.5):

$$\frac{\partial^2 S}{\partial q\,\partial t} = -\frac{\partial H}{\partial q}\left(q, \frac{\partial S}{\partial q}, t\right) - \frac{\partial H}{\partial p}\left(q, \frac{\partial S}{\partial q}, t\right)\frac{\partial^2 S}{\partial q\,\partial q} \quad .$$

Combining these two relations, we have:

$$\frac{dp(t)}{dt} = - \frac{\partial H}{\partial q}\left(q(t), \frac{\partial S}{\partial q}(t),t\right) = - \frac{\partial H}{\partial q}\left(q(t),p(t),t\right) \quad ,$$

i.e., $(q(t),p(t))$ together solve the Hamilton equations.

Now, (3.6) can be solved, with $q(t)$ prescribed at, say, $t = 0$, as any points q° of q-space. Thus, we get an n-parameter family of extremals. (However, only those n-parameter family extremals that arise in this way from a solution of the Hamilton-Jacobi equation (3.5) are called extremal fields.)

There is another interpretation of the function $S(q,t)$ that solves (3.5) in terms of <u>action</u>. (If $(q(t),p(t))$ is <u>any</u> curve in phase space, $0 \le t \le a$, then the action along the curve is

$$\int_0^a p(t) \frac{dq(t)}{dt} - H(q(t),p(t),t)dt) \quad .$$

To see this, given an extremal $q(t)$ that is imbedded in an extremal field determined by $S(q,t)$, i.e., that solves (3.6), let us compute the action along the curve $(q(t),p(t) = (\partial S/\partial q)(q(t),t))$ in phase space.

$$\text{Action} = \int_0^a p(t) \frac{dq(t)}{dt} - H(q(t),p(t),t)\, dt$$

$$= \int_0^a \frac{\partial S}{\partial q}(q(t),t) \frac{dq(t)}{dt} - H\left(q(t), \frac{\partial S}{\partial q}(q(t),t),t\right) dt$$

$$= \int_0^a \frac{\partial S}{\partial q}(q(t),t) \frac{dq(t)}{dt} + \frac{\partial S}{\partial t}(q(t),t)\, dt$$

$$= \int_0^a \frac{d}{dt}(S(q(t),t))\, dt = S(q(a),a) - S(q(0),0) \quad .$$

Thus, the action along this curve in phase space is just the difference in values of S on the end points of the corresponding extremal curve in q-space. This observation lies at the heart of the famous relation between classical mechanics and geometrical optics. The extremals of the extremal field are the "rays" corresponding to the wave fronts, which are, for fixed t, the surfaces $S(q,t) = \text{constant}$.

Suppose now that we start off with a submanifold A of q-space, with a given extremal field, with $S(q,0) = \text{constant}$ for $q \in A$. In fact, we can normalize things so that $S(q,0) = 0$ for $q \in A$. Let us choose a time interval, say $0 \leq t \leq a$. For $q^\circ \in A$, there is a unique solution of $q(t,q^\circ)$ of (3.6) with $q(0,q^\circ) = 0$. Following this along to $t = a$ gives us a mapping $q^\circ \to q(a,q^\circ)$, representing the "optical image" of the submanifold A. In geometrical optics, one is interested in actually computing this submanifold. In practice, one must be satisfied with an approximate calculation. We will now show how making this approximate calculation is equivalent to solving the linear variational equations (3.4) with an appropriate boundary condition at $t = 0$, i.e., just the sort of problem dealt with in Section 2. We can also suppose, by choosing the coordinate system correctly, that A is a linear subspace of q-space.

We can normalize so that the origin O of the q coordinates is a point of A, and so that $q(t)$ is the curve of the extremal field with $q(O) = 0$. Let s be another real parameter designed to parameterize curves along A; denote a typical curve in A by $q^\circ(s)$. We can then, for each s, find a curve $t \to q(s,t)$ of the extremal curve with $q(s,o) = q^\circ(s)$. The curve $s \to q(s,a)$ will then represent the corresponding curve in the optical image. By an "approximate calculation" we mean that we must be satisfied with computing the coefficient $Q(t)$ in the Taylor expansion:

$$q(s,t) = q(t) + Q(t)s + \dots \quad .$$

Now, for fixed s, the curve $t \to (q(s,t), p(s,t) = (\partial S/\partial q)(q(s,t),t))$ is a solution of the Hamilton equations (3.1). Thus, if the Taylor expansion of $p(s,t)$ is

$$p(s,t) = p(t) + P(t)s + \dots \quad ,$$

we have seen that $(Q(t), P(t))$ is a solution of the linear variational equations (3.4) based on the given solution $(q(t),p(t))$. To determine $(Q(t), P(t))$ completely, it only remains then to find the boundary conditions at $t = 0$. Now, $Q(0) = (\partial q/\partial s)(0,0)$; hence, $Q(0)$ must be in the tangent space to A at $q^\circ = 0$. Also, the fact that $q^\circ(s)$ lies in A and that $q \to S(q,o)$ is constant on A gives:

$$0 = \frac{d}{ds} S(q^\circ(s),0) = \frac{\partial S}{\partial q}(q^\circ(s),0) \frac{dq^\circ}{ds}$$

$$= p(s,0) \frac{dq^\circ}{ds} \quad ;$$

hence,

$$p(0)Q(0) = 0 \quad .$$

Since this must be true for essentially any curve in A starting at O, we see that the vector p(0) must be perpendicular to the tangent space to A at O. Hence we can differentiate with respect to s again and set s = 0, to obtain:

$$P(0)\, Q(0) = 0 \quad .$$

Finally, then, the boundary conditions satisfied by (Q(t), P(t)) at t = 0 are:

Q(0) belongs to the tangent space to A at O . (3.8a)

P(0) is perpendicular to the tangent space to A at O . (3.8b)

We want to transform conditions (3.8) into conditions on Q(0) and Q'(0) in order to relate to the work in Section 2. Return to the linear variational equations (3.4), written as:

$$\frac{dQ}{dt} = A(t)P(t) + B(t)Q(t) \quad ;$$

$$\frac{dP}{dt} = -\,{}^*B(t)P(t) + C(t)Q(t) \quad . \tag{3.4'}$$

(A(t), B(t), C(t)) are n x n matrices, given as follows in terms of H(p,q,t) and the base solution of (p(t),q(t)) of (3.1).

$$A_{ij}(t) = \frac{\partial^2 H}{\partial p_i \partial p_j}(q(t),p(t),t) \quad ; \quad B_{ij}(t) = \frac{\partial^2 H}{\partial p_i \partial q_j}(q(t),p(t),t) \quad ;$$

$$C_{ij}(t) = -\frac{\partial^2 H}{\partial q_i \partial q_j}(q(t),p(t),t) \quad ; \quad {}^*B_{ij}(t) = B_{ji}(t) \quad . \tag{3.9}$$

(*denotes transpose of the matrix)

We will suppose that:

$$\det A(t) \neq 0 \quad \text{for all } t \quad ; \tag{3.10}$$

i.e.,

$$\det\left(\frac{\partial^2 H}{\partial p_i \partial p_j}\right) \neq 0 \quad .$$

For the Hamiltonians arising from a calculus-of-variations problem, (3.10) corresponds to supposing that the variational problem is regular. (All of the usual problems from classical mechanics and optics satisfy this condition.) Let $A(t)^{-1}$ denote the inverse matrix. Then,

$$P = A(t)^{-1}\left(\frac{dQ}{dt} - B(t)Q(t)\right) \quad ;$$

hence:

$$\frac{d}{dt}\left(A(t)^{-1}\frac{dQ}{dt} - A(t)^{-1}B(t)Q(t)\right) = -*BA^{-1}\frac{dQ}{dt} + (*BA^{-1}B + C)Q \quad . \tag{3.11}$$

The boundary conditions satisfied at $t = 0$ now take the form:

$$Q(0) \in W \tag{3.12a}$$

$$A(0)^{-1}\left(\frac{dQ}{dt}(0) - B(0)Q(0)\right) \text{ is perpendicular to } W \quad . \tag{3.12b}$$

Of course, if A and B are suitably restricted, we immediately get a problem of the type considered in Section 2. For example, $A(t)$ = identity matrix, $B(t) = 0$ will do the job (since C is automatically a symmetric matrix). However, it is also possible to proceed directly with the second-order system of differential equations (3.11) with initial values (3.12). Morse considers this system, and the proof of Section 2 gives the Morse Index Theorem again: We dealt with the simple version in Section 2 merely to save notational energy. Note also that a simplification has been made in the initial conditions (3.12) because we chose our coordinates in q-space so that the submanifold W of q-space for which we are finding the approximate optical image is a linear subspace. Not having made this assumption would have added more terms to (3.12b), representing the "curvature" of W with respect to the q-coordinate system.

Note the physical significance of the focal point idea: A number $a < 0$ would be a focal point if there exists a nonzero solution $Q(t)$ of (3.11), satisfying (3.12) and $Q(a) = 0$. Thus, the optical image of W would degenerate in dimension at $t = a$. A designer of optical instruments would have an obvious interest in finding the first such focal point. (This is also the differential geometer's chief interest.) But, the Morse Index Theorem is ideally suited to the job of estimating the location of the first focal point. Several computations of this type can be found in Ref. (5), in the context of Riemannian geometry.

BIBLIOGRAPHY

1. N. Akhiezer, The calculus of variations, Blaisdell (1962).
2. W. Ambrose, The Cartan structural equations in classical Riemannian geometry, J. Indian Math. Society, 1960.
3. W. Ambrose, The index theorem in Riemannian geometry, Annals of Math. 73 49-86 (1961).
4. G. Birkhoff and G. C. Rota, Ordinary differential equations, Ginn and Company (1961).
5. R. Hermann, Focal points of submanifolds of Riemannian spaces, to appear in Proc. Ned. Akad. Wet.
6. E. L. Ince, Ordinary differential equations, Dover Publ., Inc., New York (1956).
7. M. Morse, Calculus of variations in the large, Am. Math. Soc. Colloquium publications (1935).
8. E. C. Whittaker, Analytical dynamics, Dover Publications, Inc., New York (1944).